

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(43) 国际公布日:

2005年8月4日(04.08.2005)

PCT

(10) 国际公布号:

WO 2005/071072 A1

- (51) 国际分类号⁷: C12N 9/12, 15/54
- (21) 国际申请号: PCT/CN2004/000039
- (22) 国际申请日: 2004年1月13日(13.01.2004)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人(对除美国以外的所有指定国): 中国科学院微生物研究所(THE INSTITUTE OF MICROBIOLOGY CHINESE ACADEMY OF SCIENCES) [CN/CN]; 中国北京市海淀区中关村北一条13号中科院微生物研究所, Beijing 100080 (CN).
- (72) 发明人; 及
- (75) 发明人/申请人(仅对美国): 高光侠(GAO, Guangxia) [CN/CN]; 刘树峰(LIU, Shufeng) [CN/CN]; 中国北京市海淀区中关村北一条13号中科院微生物研究所, Beijing 100080 (CN).

(81) 指定国(除另有指明, 要求每一种可提供的国家保护):
AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW

(84) 指定国(除另有指明, 要求每一种可提供的地区保护):
ARIPO(BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚专利(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲专利(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

本国际公布:
— 包括国际检索报告。

- (74) 代理人: 北京纪凯知识产权代理有限公司(JEEKAI & PARTNERS); 中国北京市西城区宣武门西大街甲129号金隅大厦602室, Beijing 100031 (CN).

所引用双字母代码和其它缩写符号, 请参看刊登在每期PCT公报期刊起始的“代码及缩写符号简要说明”。

(54) Title: THE RECOMBINANT MURINE LEUKEMIA VIRUS REVERSE TRANSCRIPTASE, THE GENE ENCODING IT AND THE METHOD FOR EXPRESSING IT

(54) 发明名称: 重组小鼠白血病病毒逆转录酶及其编码基因与表达方法

(57) Abstract: The present invention provides the recombinant murine leukemia virus reverse transcriptase, the gene encoding it and the method for expressing it. The MLV-RT reverse transcriptase has an amino acid sequence identical to the amino acid sequence present in wild type MLV-RT reverse transcriptase except that the amino acid at position 84 from N-terminal end is the residual X instead of Gln, in which X is a amino acid which has a shorter side chain than Gln. The recombinant murine leukemia virus reverse transcriptase of the present invention has more high enzymatic activity and continuous synthetic ability than Superscript.

(57) 摘要

本发明公开了一种重组小鼠白血病病毒逆转录酶及其编码基因与表达方法。本发明所提供的重组小鼠白血病病毒逆转录酶, 是将 MLV-RT 的自 N 端第 84 位谷氨酰胺残基取代为氨基酸残基 X 的蛋白质, 其中 X 为侧链短于所述谷氨酰胺侧链的氨基酸。本发明的重组小鼠白血病病毒逆转录酶具有比 Superscript 更高酶活性和持续合成能力, 将在生物技术领域中发挥重要作用, 可以广泛用于 cDNA 的合成。



WO 2005/071072 A1

重组小鼠白血病病毒逆转录酶及其编码基因与表达方法

技术领域

本发明涉及生物技术领域中一种重组逆转录酶及其编码基因与表达方法，特别涉及一种重组小鼠白血病病毒逆转录酶及其编码基因与表达方法。

背景技术

逆转录酶（Reverse Transcriptase，简称 RT）是由逆转录病毒编码的 DNA 聚合酶，能以 RNA 或 DNA 作为模板合成 DNA。逆转录酶是分子生物学研究的常用工具酶，可以将 RNA 逆转录成 cDNA，主要用于 cDNA 文库的构建或与聚合酶链式反应（PCR）合用（RT-PCR）检测生物样品中 RNA 的含量。目前市场上占主导地位的逆转录酶是小鼠白血病病毒（Murine Leukemia Virus，简称 MLV）的逆转录酶（MLV-RT）。

MLV-RT 由两个功能域组成，同时含有两种酶活性，即位于 N 端的 DNA 聚合酶（Pol）和位于 C 端的 RNase H。两个功能域可分别表达，而不影响任何一种酶的活性。用于 cDNA 合成的第一代重组 MLV-RT 只含有 DNA 聚合酶部分而不含有 RNaseH 部分，虽然酶的活性与全长的 MLV-RT 相似，但其持续合成能力较弱，合成的 cDNA 较短。后来人们发现，RNaseH 部分虽然不影响 MLV-RT 的 DNA 聚合酶活性，却影响其持续合成能力，原因是 RNaseH 结构域能结合模板-引物复合物，增加 MLV-RT 与模板-引物的亲和力。当利用定点突变方法将位于 RNaseH 活性中心的第 524 位的天门冬氨酸（Asp）突变为天门冬酰胺（Asn），成为 MLV-RT-D524N，该酶不具有 RNaseH 活性，但保留了其 DNA 聚合酶的活性和对模板-底物的高亲和力，因此具有更高的持续合成能力。该突变酶为美国 Invitrogen 公司专利产品，商品名为 Superscript，是目前使用最为广泛的逆转录酶。但是，Superscript 也并非完美无缺，其持续合成能力尚不理想，合成的 cDNA 中绝大部分还是不完整的。

发明公开

本发明的目的是提供一种重组小鼠白血病病毒逆转录酶。

本发明所提供的重组小鼠白血病病毒逆转录酶，名称为 MLV-RT-Q84X，是将 MLV-RT 的自 N 端第 84 位谷氨酰胺残基取代为氨基酸残基 X 的蛋白质，其中 X 为侧链短于谷氨酰胺残基侧链的氨基酸。

为了进一步提高其活性，再将上述重组小鼠白血病病毒逆转录酶的自 N

端第 524 位天门冬氨酸残基取代为天门冬酰胺残基，所得到的重组小鼠白血病病毒逆转录酶的名称为 MLV-RT-Q84X-D524N。

其中，所述 X 优选为丙氨酸、丝氨酸、天门冬氨酸或天门冬酰胺，尤其优选为丙氨酸。

5 上述重组小鼠白血病病毒逆转录酶的编码基因也属于本发明的保护范围。

本发明的另一个目的是提供一种表达重组小鼠白血病病毒逆转录酶的方法。

10 一种表达重组小鼠白血病病毒逆转录酶的方法，是将含有重组小鼠白血病病毒逆转录酶编码基因的表达载体转化到大肠杆菌中，培养阳性克隆，表达重组小鼠白血病病毒逆转录酶；所述重组小鼠白血病病毒逆转录酶，是将 MLV-RT 的自 N 端第 84 位谷氨酰胺残基取代为氨基酸残基 X 的蛋白质，其中 X 为侧链短于所述谷氨酰胺侧链的氨基酸。

15 其中，所述重组小鼠白血病病毒逆转录酶优选为将 MLV-RT 的自 N 端第 84 位谷氨酰胺残基取代为氨基酸残基 X，自 N 端第 524 位天门冬氨酸残基取代为天门冬酰胺残基的蛋白质，其中 X 为侧链短于所述谷氨酰胺侧链的氨基酸；尤其优选为将 MLV-RT 的自 N 端第 84 位谷氨酰胺残基取代为丙氨酸残基，自 N 端第 524 位天门冬氨酸残基取代为天门冬酰胺残基的蛋白质。

20 所述含有重组小鼠白血病病毒逆转录酶编码基因的表达载体可以是具有序列表中序列 1 的核苷酸序列的质粒 pTacRT-Q84A-D524N 或具有序列表中序列 3 的核苷酸序列的质粒 pTacRT-Q84N-D524N，所述大肠杆菌为 *Escherichia coli* BL21，质粒 pTacRT-Q84A-D524N 和质粒 pTacRT-Q84N-D524N 所表达的重组小鼠白血病病毒逆转录酶的氨基酸序列分别如序列表中的序列 2 和序列 4 所示。其中，序列表中的序列 1 和序列 3 均由 7488 个碱基组成，它们的开放
25 阅读框架均自 5' 端第 1515 位碱基-3527 位碱基，序列表中序列 2 和 4 均由 671 个氨基酸残基组成。

附图说明

图 1 为纯化后的 MLV-RT-Q84A-D524N SDS-PAGE 电泳检测图

图 2 为 MLV-RT-Q84A-D524N 与 MLV-RT-D524N 酶动力学分析图

30 图 3 为利用 MLV-RT-Q84A-D524N 和 MLV-RT-D524N 合成的 cDNA 的比较图

图 4 为纯化后的 MLV-RT-Q84N-D524N SDS-PAGE 电泳检测图

图 5 为 MLV-RT-Q84N-D524N 的酶活测定图

实施发明的最佳方式

实施例 1、MLV-RT-Q84A-D524N 的制备

1、质粒 pTacRT-Q84A-D524N 的构建

5 本步骤中将 MLV-RT-D524N 中的第 84 位的谷氨酰胺突变为丙氨酸，使之成为 MLV-RT-Q84A-D524N。

MLV-RT-Q84A-D524N 的定点突变采用将两段 PCR 产物 AflIII-EcoRI 和 EcoRI-MfeI 替换掉 pTacRT-D524N (*Blain, S. W. & Goff, S. P. (1995) J. Virol. 69, 4440-4452.*) 的 AflIII-MfeI 片段 (nt1467-2058)。引物 Q84A-SP (5' CGGAATTCTGGTACCCTGCCAGTC) 包含有一个由于同义突变产生的 EcoRI 酶切位点 (用下划线表示)，它和下游引物 (5' TGGGAGTCTGGTCCAGG) PCR 扩增得到一个 300bp 的 EcoRI-MfeI 片段。引物 Q84A-AP (5' CGGAATTCCC~~CG~~CGTCCAACAGTCTCTGTA) 同样包含有一个由于同义突变产生的 EcoRI 酶切位点以及突变的丙氨酸密码子 (黑体表示)，它和上游引物 (5' GTGGAATTGTGAGCCGA) 经 PCR 扩增得到一个 300bp 的 AflIII-EcoRI 片段。将
15 分别用 AflIII/EcoRI 双酶切处理过的 AflIII-EcoRI 片段 (nt1467-1770) 以及 EcoRI/MfeI 双切处理过的 EcoRI-MfeI (nt1770-2058) 片段与经 AflIII-MfeI 双切过的载体 pTacRT-D524N 6.9kb 片段连接，转化大肠杆菌 Top10。所得重组质粒 pTacRT-Q84A-D524N 通过酶切鉴定为阳性克隆，双脱氧法测序表明其具有序列表中序列 1 所示的核苷酸序列。

20 2、重组逆转录酶在大肠杆菌中的诱导和表达

将 pTacRT-Q84A-D524N 转化 *Escherichia coli* BL21，挑取单菌落接种于含 100 μ g/ml 氨苄青霉素的 LB 液体培养基中，37°C 培养至 OD₆₀₀ \approx 0.5 时，加入 IPTG 至终浓度 0.5mM，37°C 继续培养 2-3 小时，收集菌体，将菌体用预冷的 50 mM Tris-HCl (pH 7.5) 洗涤一次后备用。

25 3、重组逆转录酶的纯化

将收集的菌体悬浮于缓冲液 A (20mM PBS, pH 7.4, 0.5M NaCl)，冻融后，加入溶菌酶至终浓度 0.5mg/ml，4°C 消化 30min，超声 (功率：190W) 破碎 3 次，每次 20 秒，离心，收集上清。经过金属螯合层析柱 HiTrap chelating HP column (购自 Pharmacia) 分离纯化后，纯度即可达 80% 以上，根据 SDS-PAGE
30 显示分子量大小，确定并收集活性蛋白峰。经过离子交换层析柱 MonoS (购自 Pharmacia) 进一步分离纯化后，电泳检测考马斯亮蓝染色，结果如图 1 所示，

表明该蛋白为一分子量约为 76KD 的均一条带，未见杂质。图 1 中，M 为分子量标准；1 表示样品中含 5 微克蛋白；2 表示样品中含 2 微克蛋白。

4、酶活性测定及动力学分析

50 μ l 反应体系中包括 10ng RT 纯酶样品，60 mM Tris.Hcl (pH 8.0), 75 mM NaCl, 0.7 mM MnCl₂, 5 mM DTT, 12 μ g/ml Poly(rA) 模板, 6 μ g/ml oligo(dT)₁₈ 引物, 10 μ Ci/ml (1Ci=37 GBq) ³²P 标记的 dTTP 以及 12 μ M 未标记的 dTTP, 反应于 37°C 进行。分别于反应开始后不同时间点取样 4 μ l 点于 DE81 滤纸，终止反应。然后将 DE81 滤纸用 2 \times SSC 洗涤三次，95%乙醇漂洗两次，将滤纸晾干后，放射自显影。定量分析采用液闪计数器测定。

动力学分析采用公知的双倒数作图法进行，如图 2 所示。用每个样品中“保留在 DE81 滤纸的放射强度/同一样品总的放射强度”来表示掺入到产物中的 dTTP 的量。动力学分析显示，在以 dTTP 为底物时，MLV-RT-Q84A-D524N 和 RT-D524N 具有相似的 K_m (分别为 11.04 μ M 和 12.94 μ M)，而 MLV-RT-Q84A-D524N 的最大反应速率 V_{max} 值则是 MLV-RT-D524N 的 3.2 倍(分别为 0.41 μ mol.min⁻¹.mg⁻¹ 和 0.13 μ mol.min⁻¹.mg⁻¹) (如表 1)。

表 1. MLV-RT-Q84A-D524N 与 MLV-RT-D524N 酶动力学参数比较

逆转录酶	V _{max} (μ mol.min ⁻¹ .mg ⁻¹)	K _m (μ M)
RT-D524N	0.13 \pm 0.04	12.94 \pm 2.08
RT-Q84A-D524N	0.41 \pm 0.04	11.05 \pm 0.72

5. cDNA 合成

本步骤比较了 MLV-RT-Q84A-D524N 与 MLV-RT-D524N 在合成 cDNA 能力方面的差异。

cDNA 合成采用 293T 细胞的总 RNA 为模板，oligo(dT)₁₈ 为引物合成第一链 cDNA，产物中掺有 ³²P 作为标记。20 μ l 反应体系中包括 1 μ g 293T 细胞总 RNA, 50 mM Tris.Hcl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 500 μ M dTTP, 50 μ g/ml oligo(dT)₁₈, 2 μ Ci ³²PdTTP, 1 μ g RT 纯酶样品。42°C 反应 1 小时。碱变性电泳检测实验结果。结果如图 3 所示，表明 MLV-RT-Q84A-D524N 合成的 cDNA 片段不但较 RT-D524N 合成的长，而且合成的 cDNA 量也比 RT-D524N 多。

实施例 2、MLV-RT-Q84N-D524N 的制备

1、质粒 pTacRT-Q84N-D524N 的构建

本步骤中将 MLV-RT-D524N 中的第 84 位的谷氨酰胺突变为天门冬酰胺，使之成为 MLV-RT-Q84N-D524N。

MLV-RT-Q84N-D524N 的定点突变采用将两段 PCR 产物 AflIII-BamHI 和 BamHI-MfeI 替换掉 pTacRT-D524N (Blain, S. W. & Goff, S. P. (1995) *J. Virol.* 69, 4440-4452.) 的 AflIII-MfeI 片段 (nt1467-2058)。引物 Q84N-SP (5' CGGGATCCTGGTACCCTGCCAGTC) 包含有一个由于同义突变产生的 BamHI 酶切位点 (用下划线表示)，它和下游引物 (5' TGGGAGTCTGGTCCAGG) PCR 扩增得到一个 300bp 的 BamHI-MfeI 片段。引物 Q84N-AP (5' CGGGATCCCGTTGTCCAACAGTCTCTGTA) 同样包含有一个由于同义突变产生的 BamHI 酶切位点以及突变的天门冬酰胺密码子 (黑体表示)，它和上游引物 (5' GTGGAATTGTGAGCCGA) 经 PCR 扩增得到一个 300bp 的 AflIII-BamHI 片段。将分别用 AflIII/BamHI 双酶切处理过的 AflIII-BamHI 片段 (nt1467-1770) 以及 BamHI/MfeI 双切处理过的 BamHI-MfeI (nt1770-2058) 片段与经 AflIII-MfeI 双切过的载体 pTacRT-D524N 6.9kb 片段连接，转化大肠杆菌 Top10。所得重组质粒 pTacRT-Q84N-D524N 通过酶切鉴定为阳性克隆，双脱氧法测序表明其具有序列表中序列 3 所示的核苷酸序列。

2、重组逆转录酶在大肠杆菌中的诱导和表达

将 pTacRT-Q84N-D524N 转化 *Escherichia coli* BL21，挑取单菌落接种于含 100 μ g/ml 氨苄青霉素的 LB 液体培养基中，37°C 培养至 OD₆₀₀ \approx 0.5 时，加入 IPTG 至终浓度 0.5mM，37°C 继续培养 2-3 小时，收集菌体，将菌体用预冷的 50 mM Tris-HCl (pH 7.5) 洗涤一次后备用。

3、重组逆转录酶的纯化

将收集的菌体悬浮于缓冲液 A (20mM PBS, pH 7.4, 0.5M NaCl)，冻融后，加入溶菌酶至终浓度 0.5mg/ml，4°C 消化 30min，超声 (功率：190W) 破碎 3 次，每次 20 秒，离心，收集上清。经过金属螯合层析柱 HiTrap chelating HP column (购自 Pharmacia) 分离纯化后，纯度即可达 80% 以上，根据 SDS-PAGE 显示分子量大小，确定并收集活性蛋白峰。经过离子交换层析柱 MonoS (购自 Pharmacia) 进一步分离纯化后，电泳检测考马斯亮蓝染色，结果如图 4 所示，表明该蛋白为一分子量约为 76KD 的均一条带，未见杂质。图 4 中，M 为分子量标准；1 表示样品中含 1 微克蛋白；2 表示样品中含 2 微克蛋白；3 表示样品中含 5 微克蛋白。

4、酶活性测定及动力学分析

50 μ l 反应体系包括 10ng RT 纯酶样品, 60 mM Tris. Hcl (pH 8.0), 75 mM NaCl, 0.7 mM MnCl₂, 5 mM DTT, 12 μ g/ml Poly(rA) 模板, 6 μ g/ml oligo(dT)₁₈ 引物, 10 μ Ci/ml (1Ci=37 GBq) ³²P 标记的 dTTP 以及 12 μ M 未标记的 dTTP, 反应于 37°C 进行。分别于反应开始后不同时间点取样 4 μ l 点于 DE81 滤纸, 终止反应。然后将 DE81 滤纸用 2 \times SSC 洗涤三次, 95%乙醇漂洗两次, 将滤纸晾干后, 放射自显影。结果如图 5 所示, 图中黑点颜色的深浅代表酶活力的高低。结果显示 MLV-RT-Q84N-D524N 的酶活力较 MLV-RT-D524N 高, 跟 MLV-RT-Q84A-D524N 的活力相当。

10 工业应用

本发明通过分析 MLV-RT 的晶体结构模型, 发现该酶第 84 位的谷氨酰胺 (Q84) 位于酶的活性中心附近, 参与调控酶的催化能力, 其较长的侧链妨碍合成产物的延伸, 影响 DNA 聚合酶的活性和持续合成能力。本发明将 MLV-RT 基因进行突变, 使 Q84 残基突变成侧链较短的残基, 产生了新的突变酶 MLV-RT-Q84X, 其中 X 为侧链短于谷氨酰胺侧链的氨基酸, 如丙氨酸 (Ala)、丝氨酸 (Ser)、天门冬酰胺 (Asn)、或天门冬氨酸 (Asp) 等。同时, 本发明将 Q84X 突变引入 MLV-RT-D524N 中, 使之成为 MLV-RT-Q84X-D524N。本发明的重组小鼠白血病病毒逆转录酶具有比 Superscript 更高酶活性和持续合成能力, 将在生物技术领域中发挥重要作用, 可以广泛用于 cDNA 的合成。

权利要求书

- 1、一种重组小鼠白血病病毒逆转录酶，是将 MLV-RT 的自 N 端第 84 位谷氨酰胺残基取代为氨基酸残基 X 的蛋白质，其中 X 为侧链短于所述谷氨酰胺残基侧链的氨基酸。
- 2、根据权利要求 1 所述的重组小鼠白血病病毒逆转录酶，其特征在于：所述重组小鼠白血病病毒逆转录酶的自 N 端第 524 位天门冬氨酸残基取代为天门冬酰胺残基。
- 3、根据权利要求 1 或 2 所述的重组小鼠白血病病毒逆转录酶，其特征在于：所述 X 为丙氨酸、丝氨酸、天门冬氨酸或天门冬酰胺。
- 4、根据权利要求 3 所述的重组小鼠白血病病毒逆转录酶，其特征在于：所述 X 为丙氨酸。
- 5、编码权利要求 1 所述重组小鼠白血病病毒逆转录酶的基因。
- 6、一种表达重组小鼠白血病病毒逆转录酶的方法，是将含有重组小鼠白血病病毒逆转录酶编码基因的表达载体转化到大肠杆菌中，培养阳性克隆，表达重组小鼠白血病病毒逆转录酶；所述重组小鼠白血病病毒逆转录酶，是将 MLV-RT 的自 N 端第 84 位谷氨酰胺残基取代为氨基酸残基 X 的蛋白质，其中，X 为侧链短于所述谷氨酰胺残基侧链的氨基酸。
- 7、根据权利要求 6 所述的方法，其特征在于：所述重组小鼠白血病病毒逆转录酶的自 N 端第 524 位天门冬氨酸残基取代为天门冬酰胺残基。
- 8、根据权利要求 7 所述的方法，其特征在于：所述 X 为丙氨酸。
- 9、根据权利要求 8 所述的方法，其特征在于：所述含有重组小鼠白血病病毒逆转录酶编码基因的表达载体是具有序列表中序列 1 的核苷酸序列的质粒 pTacRT-Q84A-D524N。
- 10、根据权利要求 6、7、8 或 9 所述的方法，其特征在于：所述大肠杆菌为 *Escherichia coli* BL21。

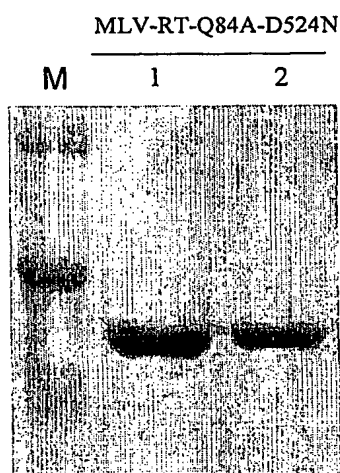


图 1

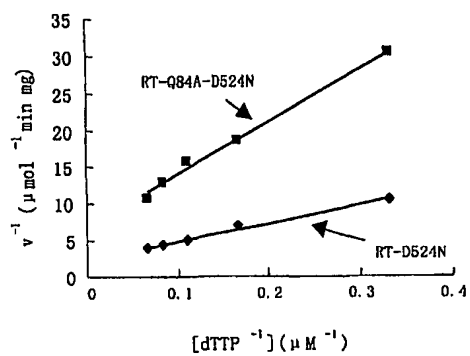


图 2

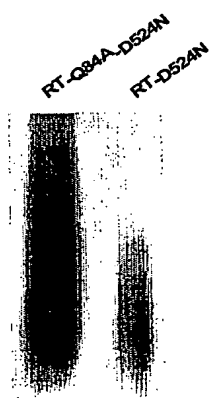


图 3

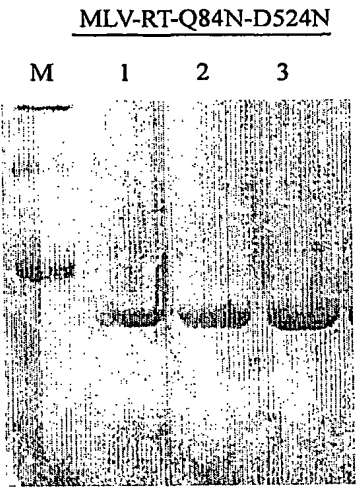


图 4

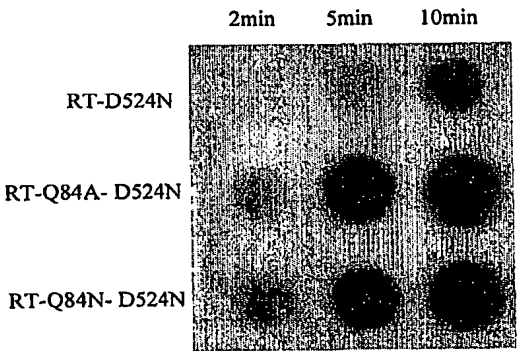


图 5

序列表

<160> 4

5 <210> 1
 <211> 7488
 <212> DNA
 <213> 人工序列

10 <220>
 <223>

<400> 1

	ccgacacccat cgaatggtgc aaaacctttc gcggtatggc atgatatgcgc ccggaagaga	60
15	gtcaattcag ggtggtgaat gtgaaaccag taacgttata cgatgtcgca gagtatgccg	120
	gtgtctctta tcagaccgtt tcccgcgtgg tgaaccaggc cagccacgtt tctgcgaaaa	180
	cgcgggaaaa agtggaaagcg gcgatggcgg agctgaatta cattcccaac cgcgtggcac	240
	aacaactggc gggcaaacag tcgttgctga ttggcgttgc cacctccagt ctggccctgc	300
	acgcgccgtc gcaaattgtc gcggcgatta aatctcgcgc cgatcaactg ggtgccagcg	360
20	tggtggtgtc gatggtagaa cgaagcggcg tcgaagcctg taaagcggcg gtgcacaatc	420
	ttctcgcgca acgcgtcagt gggctgatca ttaactatcc gctggatgac caggatgccg	480
	ttgctgtgga agctgcctgc actaatgttc cggcgttatt tcttgatgtc tctgaccaga	540
	cacccatcaa cagtattatt ttctcccatg aagacggtag gcgactgggc gtggagcatc	600
	tggtcgcatt gggtcaccag caaatcgcgc tgtagcggg ccattaagt tctgtctcgg	660
25	cggtctcgcg tctggctggc tggcataaat atctcactcg caatcaaatt cagccgatag	720
	cggaacggga aggcgactgg agtgccatgt ccggttttca acaaaccatg caaatgctga	780
	atgagggcat cgttcccact gcgatgctgg ttgccaacga tcgatggcg ctgggcgcaa	840
	tgcgcgccat taccgagtcg gggctgcgcg ttggtgcgga tatctcggtg gtgggatacg	900
	acgataccga agacagctca tgttatatcc cgccgttaac caccatcaaa caggattttc	960
30	gcctgtctggg gcaaaccagc gtggaccgct tgctgcaact ctctcagggc caggcgggtg	1020
	agggcaatca gctgttgccc gtctcactgg tgaagaaaa aaccaccctg gcgccaata	1080
	cgcaaaccgc ctctccccgc gcgttgccg attcattaat gcagctggca cgacaggttt	1140
	cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct cactcattag	1200
	gcacaattct catgtttgac agcttatcat cgactgcacg gtgcaccaat gcttctggcg	1260
35	tcaggcagcc atcggaagct gtggtatggc tgtgcaggtc gtaaatcact gcataattcg	1320
	tgctcgctcaa ggcgcactcc cggtctggat aatgtttttt gcgccgacat cataacggtt	1380
	ctggcaaata ttctgaaatg agctgttgac aattaatcat cggctcgtat aatgtgtgga	1440
	attgtgagcg gataacaatt tgaattctta agatttgtga ggggataaca atttcacaca	1500
	ggaaacagaa tatgacccta aatatagaag atgagcatcg gctacatgag acctcaaaag	1560
40	agccagatgt ttctctaggg tccacatggc tgtctgattt tcctcaggcc tgggcggaaa	1620
	ccgggggcat gggactggca gttcgccaag ctctctgat catacctctg aaagcaacct	1680
	ctacccccgt gtccataaaa caatacccca tgtcacaaga agccagactg gggatcaagc	1740
	cccacataca gagactgttg gacgcgggaa ttctggtacc ctgccagtcc ccctggaaca	1800
	cgccccgtct acccggttaag aaaccaggga ctaatgatta taggcctgtc caggatctga	1860

gagaagtcaa caagcgggtg gaagacatcc accccaccgt gcccaaccct tacaacctct 1920
 tgagcgggct cccaccgtcc caccagtggc acactgtgct tgatttaaag gatgcctttt 1980
 tctgcctgag actccacccc accagtcagc ctctcttcgc ctttgagtgg agagatccag 2040
 agatgggaat ctcaggacaa ttgacctgga ccagactccc acagggtttc aaaaacagtc 2100
 5 ccacctgtt tgatgaggca ctgcacagag acctagcaga cttccggatc cagcaccag 2160
 acttgatcct gctacagtac gtggatgact tactgtcggc cgccacttct gagctagact 2220
 gccaaacaagg tactcgggcc ctgttacaaa ccctagggaa cctcgggtat cgggcctcgg 2280
 ccaagaaagc ccaaatttgc cagaaacagg tcaagtatct ggggtatctt ctaaaagagg 2340
 gtcagagatg gctgactgag gccagaaaag agactgtgat ggggcagcct actccgaaga 2400
 10 cccctcgaca actaaggag ttccctaggga cggcaggctt ctgtcgctc tggatccctg 2460
 ggtttgcaga aatggcagcc cccttgtacc ctctcaccaa aacggggact ctgtttaatt 2520
 gggggccaga ccaacaaaag gcctatcaag aaatcaagca agctcttcta actgccccag 2580
 ccctgggggtt gccagatttg actaagccct ttgaactctt tgtcgacgag aagcagggt 2640
 acgccaagg tgctctaag caaaaactgg gaccttggcg tcggccgggt gcctacctgt 2700
 15 ccaaaaagct agaccagta gcagctgggt ggccccctg cctacggatg gtagcagcca 2760
 ttgccgtact gacaaaggat gcaggcaagc taaccatggg acagccacta gtcatcttgg 2820
 ccccccagtc agtagaggca ctagtcaaac aacccccga ccgttggtt tccaacgccc 2880
 ggatgactca ctatcaggcc ttgcttttgg acacggaccg ggtccagttc ggaccgggtg 2940
 tagccctgaa cccggctacg ctgctccac tgcctgagga agggctgcaa cacaactgcc 3000
 20 ttgatatact ggccgaagcc cacggaaccc gacccgacct aacggaccag ccgtctccag 3060
 acgccgacca cacctggtac acgaatggaa gcagtctctt acaagaggga cagcgttaagg 3120
 cgggagctgc ggtgaccacc gagaccgagg taatctgggc taaagccctg ccagccggga 3180
 catccgctca gcgggctgaa ctgatagcac tcaccaggc cctaaagatg gcagaaggta 3240
 agaagctaaa tgtttatact gatagccgtt atgcttttgc tactgccc atccatggag 3300
 25 aaatatacag aaggcgtggg ttgctcacat cagaaggcaa agagatcaaaa aataaagacg 3360
 agatcttggc cctactaaaa gccctcttct tgcccaaaag acttagcata atccattgtc 3420
 caggacatca aaagggacac agcggcgagg ctagaggcaa ccggatggct gaccaagcgg 3480
 cccgaaaggc agccatcaca gagactccag acacctctac cctcctccat caccatcacc 3540
 atcactagtc tagagtcgac ctgcaggcaa gcttggcact ggccgtcggt ttacaacgtc 3600
 30 gtgactggga aaaccctggc gttacccaac ttaatcgctt tgcagcacat ccccttttcg 3660
 ccagctggcg taatagcgaa gagggccgca ccgatcgccc ttccaacag ttgcgcagcc 3720
 tgaatggcga atggcagctt ggctgttttg gcggatgaga taagattttc agcctgatac 3780
 agattaaatc agaacgcaga agcggctctga taaaacagaa tttgcctggc ggcagtagcg 3840
 cgggtggctcc acctgacccc atgccgaact cagaagtga acgccgtagc gccgatggta 3900
 35 gtgtggggtc tccccatgcg agagtaggga actgccaggc atcaataaaa acgaaaggct 3960
 cagtcgaaag actgggcctt tcgttttata tgttgtttgt cggatgaacgc tctcctgagt 4020
 aggacaaatc cgccgggagc ggatttgaac gttgcgaagc aacggcccgg aggggtggcg 4080
 gcaggacgcc cgccataaac tgccaggcat caaattaagc agaaggccat cctgacggat 4140
 ggcctttttg cgtttctaca aactcttttt gtttattttt ctataatcat tcaaatatgt 4200
 40 atccgctcat gagacaataa cctgataaaa tgcttcaata atattgaaaa aggaagagta 4260
 tgagtattca acatttccgt gtcgccctta ttcccttttt tgcggcattt tgccttctg 4320
 tttttgctca cccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac 4380
 gagtgggtta catcgaactg gatctcaaca gcgtaagat ccttgagagt ttcgccccg 4440
 aagaacgttc tccaatgat agcactttta agttctgct atgtggcgcg gtattatccc 4500

gtgttgacgc cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg 4560
 ttgagtactc accagtcaca gaaaagcadc ttacggatgg catgacagta agagaattat 4620
 gcagtgtctgc cataacctatg agtgataaca ctgcggccaa cttacttctg acaacgatcg 4680
 gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg 4740
 5 atcgtttggga accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc 4800
 ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt 4860
 cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct 4920
 cggcccttcc ggctggctgg tttattgctg ataaatctgg agccgggtgag cgtgggtctc 4980
 gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca 5040
 10 cgacgggggag tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct 5100
 cactgattaa gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt 5160
 taccocgggt gataatcaga aaagcccaa aaacaggaag attgtataag caaatattta 5220
 aattgtaaac gttaatattt tgttaaaatt cgcgttaaat ttttgttaaa tcagctcatt 5280
 ttttaaccaa taggccgaaa tcggcaaaat cccttataaa tcaaagaat agcccgagat 5340
 15 agggttgagt gttgttccag tttggaacaa gagtccacta ttaaagaacg tggactccaa 5400
 cgtcaaaggg cgaaaaaccg tctatcaggg cgatggccca ctacgtgaac catcacccaa 5460
 atcaagtttt ttggggctga ggtgccgtaa agcactaaat cggaacccta aaggagagccc 5520
 ccgatttaga gcttgacggg gaaagccggc gaacgtggcg agaaaggaag ggaagaaagc 5580
 gaaaggagcg ggcgctaggg cgctggcaag tgtagcggtc acgctgcgag taaccaccac 5640
 20 acccgccgag cttaatgcgc cgctacaggg cgcgtaaaag gatctaggtg aagatccttt 5700
 ttgataatct catgaccaa atcccttaac gtgagttttc gttccactga gcgtcagacc 5760
 ccgtagaaaa gatcaaagga tcttcttgag atcctttttt tctgcgcgta atctgctgct 5820
 tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgggt gccggatcaa gagctaccaa 5880
 ctctttttcc gaaggtaact ggcttcagca gagcgcagat accaaatact gtccttctag 5940
 25 tgtagccgta gttaggccac cacttcaaga actctgtagc accgcctaca tacctcgctc 6000
 tgctaatect gttaccagt gctgctgcca gtggcgataa gtcgtgtctt accgggttgg 6060
 actcaagacg atagttaccg gataaggcgc agcggctcgg ctgaacgggg ggttcgtgca 6120
 cacagcccag cttggagcga acgacctaca ccgaactgag atacctacag cgtgagctat 6180
 gagaaagcgc cacgcttccc gaaggagaga aggcggacag gtatccggtg agcggcaggg 6240
 30 tcggaacagg agagcgcacg agggagcttc cagggggaaa cgcctgggtat ctttatagtc 6300
 ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt gtgatgctc tcaggggggc 6360
 ggagcctatg gaaaaacgcc agcaacgcgg cctttttacg gttcctggcc ttttgctggc 6420
 cttttgctca catgttcttt cctgcgttat cccctgattc tgttgataac cgtattaccg 6480
 cctttgagtg agctgatacc gctcgcgcga gccgaacgac cgagcgcagc gagtcagtga 6540
 35 gcgaggaagc ggaagagcgc ctgatgcggt attttctcct tacgcatctg tgcggtattt 6600
 cacaccgcat atggtgcact ctcatgataa tctgctctga tgccgcatag ttaagccagt 6660
 atacactccg ctatcgctac gtgactgggt catggctgag ccccgacacc cgccaacacc 6720
 cgctgacgag ccctgacggg cttgtctgct cccggcatcc gcttacagac aagctgtgac 6780
 cgtctccggg agctgcatgt gtcagaggtt ttcaccgtca tcaccgaaac gcgcgaggca 6840
 40 gctgcggtaa agctcatcag cgtggtcgtg cagcgattca cagatgtctg cctgttcac 6900
 cgcgtccagc tcgttgagtt tctccagaag cgtaaatgtc tggcttctga taaagcgggc 6960
 catgttaagg gcggtttttt cctgtttggg cacttgatgc ctccgtgtaa gggggaattt 7020
 ctgttcatgg gggtaatgat accgatgaaa cgagagagga tgctcacgat acgggttact 7080
 gatgatgaac atgcccgggt actggaacgt tgtgagggta aacaactggc ggtatggatg 7140

cggcgggacc agagaaaaat cactcagggt caatgccagc gcttcgttaa tacagatgta 7200
 ggtgttccac agggtagcca gcagcatcct gcgatgcaga tccggaacat aatgggtgcag 7260
 ggcgctgact tccgcgtttc cagactttac gaaacacgga aaccgaagac cattcatgtt 7320
 gttgctcagg tcgcagacgt tttgcagcag cagtcgcttc acgttcgctc gcgtatcggt 7380
 5 gattcattct gctaaccagt aaggcaaccc cgccagccta gccgggtcct caacgacagg 7440
 agcacgatca tgcgcacccg tggccaggac ccaacgctgc ccgaaatt 7488

<210> 2
 <211> 671
 10 <212> PRT
 <213> 人工序列

<220>
 <223>

15 <400> 2
 Thr Leu Asn Ile Glu Asp Glu His Arg Leu His Glu Thr Ser Lys Glu
 1 5 10 15
 Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
 20 20 25 30
 Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
 35 40 45
 Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
 50 55 60
 25 Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
 65 70 75 80
 Leu Leu Asp Ala Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
 85 90 95
 Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
 100 105 110
 30 Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
 115 120 125
 Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
 130 135 140
 35 Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
 145 150 155 160
 His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
 165 170 175
 Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
 180 185 190
 40 Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala
 195 200 205
 Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
 210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
 225 230 235 240
 Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
 245 250 255
 5 Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
 260 265 270
 Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
 275 280 285
 Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
 10 290 295 300
 Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
 305 310 315 320
 Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp
 325 330 335
 15 Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
 340 345 350
 Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
 355 360 365
 Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
 20 370 375 380
 Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
 385 390 395 400
 Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
 405 410 415
 25 Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
 420 425 430
 Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
 435 440 445
 Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
 30 450 455 460
 Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
 465 470 475 480
 Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
 485 490 495
 35 Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
 500 505 510
 Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asn Gly Ser Ser Leu
 515 520 525
 Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
 40 530 535 540
 Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
 545 550 555 560
 Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
 565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
 580 585 590
 Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
 595 600 605
 5 Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
 610 615 620
 Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
 625 630 635 640
 Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
 10 645 650 655
 Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu
 660 665 670

<210> 3
 15 <211> 7488
 <212> DNA
 <213> 人工序列

<220>
 20 <223>

<400> 3
 ccgacacccat cgaatggtgc aaaacctttc gcggtatggc atgatatgcgc ccggaagaga 60
 gtcaattcag ggtggtgaat gtgaaaccag taacgttata cgatgtcgca gagtatgccg 120
 25 gtgtctctta tcagaccgtt tcccgctgg tgaaccaggc cagccacgtt tctgcgaaaa 180
 cgcgggaaaa agtgaagcg gcgatggcgg agctgaatta cattcccaac cgcgtggcac 240
 aacaactggc gggcaaacag tcgttgctga ttggcgttgc cacctccagt ctggccctgc 300
 acgcgcgcgc gcaaattgtc gcggcgatta aatctcgcgc cgatcaactg ggtgccagcg 360
 tgggtggtgtc gatggtagaa cgaagcggcg tcgaagcctg taaagcggcg gtgcacaatc 420
 30 ttctcgcgca acgcgtcagt gggctgatca ttaactatcc gctggatgac caggatgcca 480
 ttgctgtgga agctgcctgc actaatgttc cggcgttatt tcttgatgtc tctgaccaga 540
 caccatcaa cagtattatt ttctcccatg aagacggtag gcgactgggc gtggagcatc 600
 tggtcgcatt gggtcaccag caaatcgcgc tgtagcggg ccattaagt tctgtctcgg 660
 cgcgctctgcg tctggctggc tggcataaat atctcactcg caatcaaatt cagccgatag 720
 35 cggaacggga aggcgactgg agtgccatgt ccggttttca acaaaccatg caaatgtctga 780
 atgagggcat cgttcccact gcgatgctgg ttgccaacga tcagatggcg ctgggcgcaa 840
 tgcgcgccat taccgagtcg gggctgcgcg ttggtgcgga tatctcggtg gtgggatacg 900
 acgataccga agacagctca tggtatatcc cgccgttaac caccatcaaa caggattttc 960
 gcctgctggg gcaaaccagc gtggaccgct tgctgcaact ctctcagggc caggcgggtga 1020
 40 agggcaatca gctgttgcgc gtctcactgg tgaaaagaaa aaccaccctg gcgccaata 1080
 cgcaaaccgc ctctccccgc gcgttggccg attcattaat gcagctggca cgacagggtt 1140
 cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct cactcattag 1200
 gcacaattct catgtttgac agcttatcat cgactgcacg gtgcaccaat gcttctggcg 1260
 tcaggcagcc atcggaagct gtggtatggc tgtgcaggtc gtaaatact gcataattcg 1320

	tgctgctcaa ggcgcactcc cgctctggat aatgtttttt gcgccgacat cataacggtt	1380
	ctggcaaata ttctgaaatg agctgttgac aattaatcat cggctcgtat aatgtgtgga	1440
	attgtgagcg gataacaatt tgaattctta agattttgtga ggggataaca atttcacaca	1500
	ggaaacagaa tatgacccta aatatagaag atgagcatcg gctacatgag acctcaaaag	1560
5	agccagatgt ttctctaggg tccacatggc tgtctgattt tcctcaggcc tgggcggaaa	1620
	ccgggggcat gggactggca gttcgccaag ctctctgcat catacctctg aaagcaacct	1680
	ctacccccgt gtccataaaa caatacccca tgtcacaaga agccagactg gggatcaagc	1740
	cccacataca gagactgttg gacaacggga tcctgggtacc ctgccagtcc ccctggaaca	1800
	cgcccctgct acccgtaaag aaaccaggga ctaatgatta taggcctgtc caggatctga	1860
10	gagaagtcaa caagcgggtg gaagacatcc accccacctg gccaacctt tacaacctct	1920
	tgagcgggct cccacctgcc caccagtggc acactgtgct tgatttaaag gatgcctttt	1980
	tctgcctgag actccacccc accagtcagc ctctcttcgc ctttgagtgg agagatccag	2040
	agatgggaat ctgaggacaa ttgacctgga ccagactccc acagggtttc aaaaacagtc	2100
	ccacctgtt tgatgaggca ctgcacagag acctagcaga cttccggatc cagcaccag	2160
15	acttgatcct gctacagtac gtggatgact tactgtctggc cgccacttct gagctagact	2220
	gccaacaagg tactcggggc ctgttacaaa ccctagggaa cctcgggtat cgggcctcgg	2280
	ccaagaaagc ccaaatttgc cagaaacagg tcaagtatct ggggtatctt ctaaaagagg	2340
	gtcagagatg gctgactgag gccagaaaag agactgtgat ggggcagcct actccgaaga	2400
	cccctcgaca actaaggag ttcttaggga cggcaggctt ctgtcgctc tggatccctg	2460
20	ggtttgcaga aatggcagcc cccttgtacc ctctcacaa aacggggact ctgtttaatt	2520
	ggggcccaga ccaacaaaag gcctatcaag aaatcaagca agctcttcta actgcccag	2580
	ccctggggtt gccagatttg actaagccct ttgaactctt tgtcgacgag aagcagggt	2640
	acgccaagg tgctctaacg caaaaactgg gaccttggcg tcggccgggtg gcctacctgt	2700
	ccaaaaagct agaccagta gcagctgggt ggcccccttg cctacggatg gtagcagcca	2760
25	ttgccgtact gacaaaggat gcaggcaagc taaccatggg acagccacta gtcattcttg	2820
	cccccatgc agtagaggca ctagtcaaac aacccccga ccgctggctt tccaacgccc	2880
	ggatgactca ctatcaggcc ttgcttttgg acacggaccg ggtccagttc ggaccggtg	2940
	tagccctgaa cccggctacg ctgctccac tgcttagga agggctgcaa cacaactgcc	3000
	ttgatatcct ggccgaagcc cacggaacct gacctgacct aacggaccag ccgctccag	3060
30	acgccgacca cacctggtac acgaatggaa gcagtctctt acaagaggga cagcgtaagg	3120
	cgggagctgc ggtgaccacc gagaccgagg taatctgggc taaagccctg ccagccggga	3180
	catccgctca gcgggctgaa ctgatatcac tcaccaggc cctaaagatg gcagaaggta	3240
	agaagctaaa tgtttatact gatagccgtt atgcttttgc tactgcccac atccatggag	3300
	aaatatacag aaggcgtggg ttgctcacat cagaaggcaa agagatcaaa aataaagacg	3360
35	agatcttggc cctactaaaa gccctctttc tgcccaaaag acttagcata atccattgtc	3420
	caggacatca aaaggacac agcgccgagg ctagaggcaa ccggatggct gaccaagcgg	3480
	cccgaaggc agccatcaca gagactccag acacctctac cctcctccat caccatcacc	3540
	atcactagtc tagagtcgac ctgcaggcaa gcttggcact ggccgtcgtt ttacaacgtc	3600
	gtgactggga aaaccctggc gttacccaac ttaatcgcc tgcagcacat ccccttttcg	3660
40	ccagctggcg taatagcgaa gagggccgca ccgctcgccc ttcccaacag ttgcgcagcc	3720
	tgaatggcga atggcagctt ggctgttttg gcggatgaga taagattttc agcctgatac	3780
	agattaaatc agaacgcaga agcggctctga taaaacagaa tttgcctggc ggcagtagcg	3840
	cgggtggctcc acctgacccc atgccgaact cagaagtga acgccgtagc gccgatggta	3900
	gtgtggggtc tcccatgctg agagtaggga actgccaggc atcaataaaa acgaaaggct	3960

	cagtcgaaag actgggcctt tcgttttata tgttgttgt cgggtgaacgc tctcctgagt	4020
	aggacaaatc cgccgggagc ggatttgaac gttgcgaagc aacggcccgg aggggtggcgg	4080
	gcaggacgcc cgccataaac tgccaggcat caaattaagc agaaggccat cctgacggat	4140
	ggcctttttg cgtttctaca aactcttttt gtttattttt ctaaatacat tcaaatatgt	4200
5	atccgctcat gagacaataa ccctgataaa tgcttcaata atattgaaaa aggaagagta	4260
	tgagtattca acatttccgt gtcgccctta ttcccttttt tgcggcattt tgccttcctg	4320
	tttttgctca cccagaaacg ctgggtgaaag taaaagatgc tgaagatcag ttgggtgcac	4380
	gagtgggtta catcgaactg gatctcaaca gcggtgaagat ccttgagagt ttctgccccg	4440
	aagaacgttc tccaatgatg agcactttta aagttctgct atgtggcgcg gtattatccc	4500
10	gtgttgacgc cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg	4560
	ttgagtactc accagtcaca gaaaagcatc ttacggatgg catgacagta agagaattat	4620
	gcagtgctgc cataaccatg agtgataaca ctgcggccaa cttacttctg acaacgatcg	4680
	gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg	4740
	atcgttggga accggagctg aatgaagcca tacciaacga cgagcgtgac accacgatgc	4800
15	ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt	4860
	cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct	4920
	cggcccttcc ggctggctgg tttattgctg ataaatctgg agccgggtgag cgtgggtctc	4980
	gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca	5040
	cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgtctgag ataggtcct	5100
20	cactgattaa gcattggtta ctgtcagacc aagtttactc atatatactt tagattgatt	5160
	taccccggtt gataatcaga aaagcccaa aaacaggaag attgtataag caaatattta	5220
	aattgtaaac gttaatattt tgttaaaatt cgcgttaaat tttgttaaa tcagctcatt	5280
	ttttaaccaa taggccgaaa tcggcaaaat cccttataaa tcaaaagaat agcccgagat	5340
	agggttgagt gttgttccag tttggaacaa gagtccacta ttaaagaacg tggactccaa	5400
25	cgtcaaaggg cgaaaaaccg tctatcaggg cgatggccca ctacgtgaac catcaccaa	5460
	atcaagtttt ttggggctga ggtgccgtaa agcactaaat cggaacccta aaggagagccc	5520
	ccgatttaga gcttgacggg gaaagccggc gaacgtggcg agaaaggag ggaagaaagc	5580
	gaaaggagcg ggcgctaggg cgctggcaag tgtagcggtc acgctgcgcg taaccaccac	5640
	acccgcccgc cttaatgcgc cgctacaggg cgcgtaaaag gatctaggtg aagatccttt	5700
30	ttgataatct catgaccaa atcccttaac gtgagttttc gttccactga gcgtcagacc	5760
	ccgtagaaaa gatcaaagga tcttcttgag atcctttttt tctgcgcgta atctgctgct	5820
	tgcaaacaaa aaaaccaccg ctaccagcgg tggtttggtt gccggatcaa gagctaccaa	5880
	ctctttttcc gaaggtaact ggcttcagca gagcgcagat accaaatact gtccttctag	5940
	tgtagccgta gttaggccac cacttcaaga actctgtagc accgcctaca tacctcgctc	6000
35	tgctaatacct gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttgg	6060
	actcaagacg atagttaccg gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca	6120
	cacagcccag cttggagcga acgacctaca ccgaactgag atacctacag cgtgagctat	6180
	gagaaagcgc cacgcttccc gaaggagaaa agggcgacag gtatccggta agcggcaggg	6240
	tcggaacagg agagcgcacg agggagcttc cagggggaaa cgcctgggtat ctttatagtc	6300
40	ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt gtgatgctcg tcaggggggc	6360
	ggagcctatg gaaaaacgcc agcaacggc cttttttacg gttcctggcc ttttgctggc	6420
	cttttgctca catgttcttt cctgcgttat cccttgattc tgttgataac cgtattaccg	6480
	cccttgagtg agctgatacc gctcgcgcga gccgaacgac cgagcgcagc gagtcagtga	6540
	gcgaggaagc ggaagagcgc ctgatgcggt attttctcct tacgcatctg tgcggtattt	6600

cacaccgcac atggtgcaact ctacgtacaa tctgctctga tgccgcatag ttaagccagt 6660
 atacactccg ctatcgctac gtgactgggt catggctgcg ccccgacacc cgccaacacc 6720
 cgctgacgcg cctgacggg cttgtctgct cccggcatcc gcttacagac aagctgtgac 6780
 cgtctccggg agctgcatgt gtcagagggt ttaccgtca tcaccgaaac gcgcgaggca 6840
 5 gctgcggtaa agctcatcag cgtggctcgtg cagcgattca cagatgtctg cctgttcac 6900
 cgcgctccagc tcgttgagtt tctccagaag cgtaaatgtc tggcttctga taaagcgggc 6960
 catgttaagg gcggtttttt cctgttttgt cacttgatgc ctccgtgtaa gggggaattt 7020
 ctgttcacgg gggtaatgat accgatgaaa cgagagagga tgctcacgat acgggttact 7080
 gatgatgaac atgcccgggt actggaacgt tgtgagggtg aacaactggc ggtatggatg 7140
 10 cggcggggacc agagaaaaat cactcagggt caatgccagc gcttcgttaa tacagatgta 7200
 ggtgttccac agggtagcca gcagcatcct gcgatgcaga tccggaacat aatggtgcag 7260
 ggcgctgact tccgcgtttc cagactttac gaaacacgga aaccgaagac cattcatgtt 7320
 gttgctcagg tcgcagacgt tttgcagcag cagtcgcttc acgttcgctc gcgtatcggg 7380
 gattcattct gctaaccagt aaggcaaccc cgccagccta gccgggtcct caacgacagg 7440
 15 agcacgatca tgccgacccg tggccaggac ccaacgctgc ccgaaatt 7488

<210> 4

<211> 671

<212> PRT

20 <213> 人工序列

<220>

<223>

25 <400> 4

Thr Leu Asn Ile Glu Asp Glu His Arg Leu His Glu Thr Ser Lys Glu
 1 5 10 15
 Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
 20 25 30
 30 Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
 35 40 45
 Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
 50 55 60
 35 Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
 65 70 75 80
 Leu Leu Asp Asn Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
 85 90 95
 Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
 100 105 110
 40 Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
 115 120 125
 Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
 130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
 145 150 155 160
 His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
 165 170 175
 5 Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
 180 185 190
 Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala
 195 200 205
 Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
 210 215 220
 10 Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
 225 230 235 240
 Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
 245 250 255
 15 Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
 260 265 270
 Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
 275 280 285
 Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
 290 295 300
 20 Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
 305 310 315 320
 Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp
 325 330 335
 25 Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
 340 345 350
 Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
 355 360 365
 Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
 370 375 380
 30 Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
 385 390 395 400
 Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
 405 410 415
 35 Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
 420 425 430
 Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
 435 440 445
 Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
 450 455 460
 40 Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
 465 470 475 480
 Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
 485 490 495

	Asp	Ile	Leu	Ala	Glu	Ala	His	Gly	Thr	Arg	Pro	Asp	Leu	Thr	Asp	Gln
					500				505					510		
	Pro	Leu	Pro	Asp	Ala	Asp	His	Thr	Trp	Tyr	Thr	Asn	Gly	Ser	Ser	Leu
			515					520					525			
5	Leu	Gln	Glu	Gly	Gln	Arg	Lys	Ala	Gly	Ala	Ala	Val	Thr	Thr	Glu	Thr
		530					535					540				
	Glu	Val	Ile	Trp	Ala	Lys	Ala	Leu	Pro	Ala	Gly	Thr	Ser	Ala	Gln	Arg
	545					550					555				560	
	Ala	Glu	Leu	Ile	Ala	Leu	Thr	Gln	Ala	Leu	Lys	Met	Ala	Glu	Gly	Lys
10					565				570						575	
	Lys	Leu	Asn	Val	Tyr	Thr	Asp	Ser	Arg	Tyr	Ala	Phe	Ala	Thr	Ala	His
			580						585					590		
	Ile	His	Gly	Glu	Ile	Tyr	Arg	Arg	Arg	Gly	Leu	Leu	Thr	Ser	Glu	Gly
		595						600					605			
15	Lys	Glu	Ile	Lys	Asn	Lys	Asp	Glu	Ile	Leu	Ala	Leu	Leu	Lys	Ala	Leu
		610					615					620				
	Phe	Leu	Pro	Lys	Arg	Leu	Ser	Ile	Ile	His	Cys	Pro	Gly	His	Gln	Lys
	625					630					635					640
	Gly	His	Ser	Ala	Glu	Ala	Arg	Gly	Asn	Arg	Met	Ala	Asp	Gln	Ala	Ala
					645				650					655		
20	Arg	Lys	Ala	Ala	Ile	Thr	Glu	Thr	Pro	Asp	Thr	Ser	Thr	Leu	Leu	
				660					665					670		

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2004/000039

A. CLASSIFICATION OF SUBJECT MATTER

IPC⁷, C12N9/12, C12N15/54

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC⁷, C12N9/12, C12N15/54

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI (murine and leukemia and virus and reverse and transcriptase) ;

CPRS(小鼠白血病病毒逆转录酶), Genbank;

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US, A, 6136582 (UNIV COLUMBIA NEW YORK) 24. October 2000, See the whole document	1—10
A	US, A1, 2003021799 (HUNT C L et al) 30. January 2003, See the whole document	1—10

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
01. September 2004(01.09.04)

Date of mailing of the international search report

16 · SEP 2004 (16 · 09 · 2004)

Name and mailing address of the ISA/

The Chinese Patent Office
6, Xitucheng Road, Haidian District, Beijing, 100088, China
Facsimile No. (86-10)62019451

Authorized officer

PAN, Aiqun

Telephone No. (86-10)62085349



INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2004/000039

Patent document cited in search report	Publication date	Patent family members	Publication date
US—A—6136582	24-10-2000	None	
US—A1—2003021799	30-01-2003	None	

国际检索报告

国际申请号
PCT/CN2004/000039

A. 主题的分类

IPC⁷, C12N9/12, C12N15/54

按照国际专利分类表(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC⁷, C12N9/12, C12N15/54

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

WPI (murine and leukemia and virus and reverse and transcriptase);

CPRS(小鼠白血病病毒逆转录酶), Genbank;

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	US, A, 6136582 (纽约哥伦比亚大学) 24.10月2000, 参见全文	1—10
A	US, A1, 2003021799 (HUNT C L等) 30.1月2003, 参见全文	1—10

☐ 其余文件在 C 栏的续页中列出。

☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期
01.9月2004(01.09.04)

国际检索报告邮寄日期
16·9月2004(16·09·2004)

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区衙门桥西土城路6号100088
传真号: (86-10)62019451

授权官员

潘爱群

电话号码: (86-10)62085349



国际申请号
PCT/CN2004/000039

检索报告中引用的 专利文件	公布日期	同族专利	公布 日期
US—A—6136582	24-10-2000	无	
US—A1—2003021799	30-01-2003	无	